ISCTE Business School
Instituto Universitário de Lisboa

OWL ONTOLOGY QUALITY ASSESSMENT AND OPTIMIZATION IN THE CYBERSECURITY DOMAIN

Ana Margarida Conceição Pipa

Dissertation submitted as partial requirement for the conferral of

Master in Business Management

Supervisor:

Prof. Dr. Vítor Basto Fernandes, Assistant Professor ISCTE, School of Technology and Architecture - Department of Information Science and Technology

Co-supervisor:
Prof. Dr. Iryna Yevseyeva, Senior Lecturer, De Montfort University, School of Computer Science and Informatics

October 2018

Acknowledgement

I wish to acknowledge the help of the Department of Information Science and Technology of ISCTE-IUL for the opportunities granted through the participation at conferences and the cooperation of all who participated in this study.

I am grateful to Professor Vítor Basto Fernandes of ISCTE-IUL for all the brilliant assistance and support at various stages of the research and to Professor Iryna Yevseyeva of De Montfort University Leicester for her clarification. An extended thanks to both for being open-minded and flexible regarding communication and physical and online meetings. I appreciate their dedication, advice and encouragement. Their vital inputs and guidance have been essential for the development of this study. I am also indebted to Professor Pedro, Miles and Luis for their assistances with the analysis and for the preparation of this study.

My husband and my family have played an integral role in my graduate studies. I am truly grateful for all their support and encouragement.

Reprint requests and correspondence concerning this thesis should be addressed to: Ana M.C. Pipa, mailto: a26234@iscte-iul.pt

“The mind that opens to a new idea ever it will come back in your original size.”

Einstein

Abstract

The purpose of this dissertation is to assess the quality of ontologies in patterns perceived by cybersecurity context. A content analysis between ontologies indicated that there were more pronounced differences in OWL ontologies in the cybersecurity field. Results showed an increase of relevance from expressivity to variability. Additionally, no differences were found in strategies used in most of the incidents. The ontology background needs to be emphasized to understand the quality of the phenomena. In addition, ontologies are a means of representing an area of knowledge through their semantic structure. The search of information and integration of data from different origins provides a common base that guarantees the coherence of the data. This can be categorized and described in a normative way. The unification of information with the world that surrounds us allows to create synergies between entities and relationships. However, the area of cybersecurity is one of the real-world domains where knowledge is uncertain. It is therefore necessary to analyze the challenges of choosing the appropriate representation of un-structured information. Vulnerabilities are identified, but incident response is not an automatic mechanism for understanding and processing unstructured text found on the web.

Keywords: OWL Ontology, Web Semantic, ISO, Quality metrics for ontologies, Cybersecurity Ontologies

JEL classification system

C61 - Optimization Techniques

C63 - Computational Techniques

Resumo

O objetivo desta dissertação foi avaliar a qualidade das ontologias, em padrões percebidos pelo contexto de cibersegurança. Uma análise de conteúdo entre ontologias indicou que havia diferenças mais pronunciadas por ontologias OWL no campo da cibersegurança. Os resultados mostram um aumento da relevância de expressividade para a variabilidade. Além disso, não foram encontradas diferenças em estratégias utilizadas na maioria dos incidentes. O conhecimento das ontologias precisa de ser enfatizado para se entender os fenómenos de qualidade. Além disso, as ontologias são um meio de representar uma área de conhecimento através da sua estrutura semântica e facilita a pesquisa de informações e a integração de dados de diferentes origens, pois fornecem uma base comum que garante a coerência dos dados, categorizados e descritos, de forma normativa. A unificação da informação com o mundo que nos rodeia permite criar sinergias entre entidades e relacionamentos. No entanto, a área de cibersegurança é um dos domínios do mundo real em que o conhecimento é incerto e é fundamental analisar os desafios de escolher a representação apropriada de informações não estruturadas. As vulnerabilidades são identificadas, mas a resposta a incidentes não é um mecanismo automático para se entender e processar textos não estruturados encontrados na web.

Palavras-chave: Ontologia OWL, Web Semântica, ISO, Métricas de qualidade para ontologias, Ontologias de cibersegurança

Sistema de Classificação JEL

C61 – Técnicas de otimização

C63 – Técnicas computacionais

Index

Index of tables

List of figures

Abbreviations

API	Application Programming Interface
DL	Description Logics
ISO	International Organization for Standardization
HTML	Hypertext Markup Language
MITRE	The MITRE Corporation
NIST	National Institute of Standards and Technology
OWL	Ontology Web Language
RDF	Resource Description Framework
SWRL	Semantic Web Rule Language
UML	Unified Modeling Language
WEB	World Wide Web
W3C	World Wide Web Consortium
XML	Extensible Markup Language

1. Introduction

The main objective of this chapter is to make an introduction about the area of work and to explain the research relevance of this theme. For this, the concerns that result from the limitations of existing technologies are explained. Based on these limitations, a set of objectives was defined.

The increasing use of technology has led organizations to a progressive digital transformation of their business processes. The greater dependence on technologies carries a greater risk in scenarios of service discontinuity. Thus, there is an urgent need to create contingencies to respond to possible disruptions in the systems. Currently there is no standards-based, best practice framework or scientifically validated solution that addresses the specific needs of organizations to cope with the increasing cybersecurity risks. Professionals and researchers have traditionally addressed the problems of identifying, capturing and representing knowledge in this area by non-standard means of a domain in information systems.

Considering the problem described above, in order to build ontologies, the application environment needs to be meticulous. To address the difficult described above, it is crucial to determine the scope and to define and organize in a systematic and standardized way the concepts of a domain into taxonomy, attributes, relations, instances, axioms and functions among other variables that model the domain. None of the existing research works in this areas are fully mature and each group applies their own approach to the proposals. For Morais (2013) there is a lack of ontological commitment to represent vocabulary and structures that allow an effective semantic analysis to be done by reducing the semantic ambiguity of natural language. The existence of different interpretations between business/entities and between professionals/individuals can produce conceptual misalignment and severe misunderstandings.

It is known that ontology - the philosophical study of being, is of great help to define the basic categories of being and their relations, dealing with questions concerning what entities exist, how they may be grouped, related within a hierarchy, and subdivided according to similarities and differences [1].

Ontologies are constructed in different paradigms, in which each language has its own syntax, expressiveness, reasoning ability and specificity of models. In a globalized, highly connected world, ontologies help on setting the grounds for people/machines, data, information and knowledge sharing. For Barchini, Álvarez, & Herrera a large part of the technology community is unaware that ontologies can help build better and interoperable information systems [2].

1.1. Motivation

This thesis aims to explore the ways knowledge can be represented, stored, processed and shared, to promote a common understanding of a domain, with special interest and focus in the recent area of cybersecurity.

The study focuses on information as an important resource in organizations and in cybersecurity as the means to protect this valuable asset of organizations. The companies competitiveness and success is linked to the value that they give to information. This covers people, machines and methods organized to collect, process, transmit and disseminate data. Information systems are essentially assets that capture and represent knowledge about certain domains. The growing amount of information available in the digital environment, especially on the web, increases the importance of secure systems with relevant and reliable information.

The semantic web, a set of standards and technologies produced by regulatory and standardization bodies and information and communication technologies industry players, allows to treat the semantics of the available data in the web. Ontologies, more specifically Ontology Web Language (OWL) ontologies constitute one of the main building blocks of the semantic web.

This area has been chosen for research due to the possibility of surpassing results intended for the human being, using a language capable of formally dealing with information in a context favorable to machines.

In addition, ontologies are a means of representing an area of knowledge through its semantic structure. They can facilitate the search of information and integration of data from different origins. This is because they provide a common base of understanding that guarantees the coherence of the data, which is categorized and described, in a standard way. The collection and integration of information that surrounds us, allows us to create synergies and value by knowledge discovery from entities and their relationships. However, the area of cybersecurity is one of the real-world domains where knowledge production is recent and still unstable. It is necessary to analyze the challenges of choosing the appropriate vocabulary, representation of (old and new) concepts, relations, domain rules, etc. Concepts such as vulnerability, incident, incident response, etc., need to be understood in a harmonized way, by a large and heterogeneous community of cyber players. Computers and software users, software developers, project managers and top managers, security officers, law forces, etc., need to have a shared, common understanding of cybersecurity domain concepts in order to prevent, detect, and respond to information infrastructure threats, accidents and intentional attacks.

1.2. Research problem

The increasing need of representing knowledge in a systematic and standard way for people and machines to share and exchange knowledge, lead to the raise of ontologies production in all areas of human activities in the last decade, including in the cybersecurity domain.

The huge amount of ontologies available, create a decision problem of ontology selection, when we are searching for an ontology in one of our domains of interest. Ontologies can be used for a variety of purposes, such as allow someone for fast learning of a domain of knowledge, machines to understand the context they are operating, machines to behave and interact at semantic levels, being able to deal with the meaning of external stimulus, etc. The value and utility of ontologies for the mentioned purposes, makes them attractive enough for someone to search for existing/available ontologies in a domain of knowledge, before trying to build them from scratch.

Since the ontologies of potential interest for us, might be too many for us to be able to inspect them manually, some methodology, metrics and tools must be developed to support the ontology selection process, according to some ontology user preferences and needs.

1.3. Research questions and objectives

This thesis proposes a contribution in the area of ontology quality assessment and optimization, in particular, OWL ontology quality assessment and optimization. It intends to answer totally or partially the following research questions:

- What are currently the metrics defined for ontology quality assessment?
- Is there a subset of quantitative metrics allowing for automatic ontology quality assessment computation in the cybersecurity domain?
- Is it possible to optimize ontologies that have a computer-base standard representation?

The main objective of this work is to identify a set of relevant ontology quality metrics, to be used in ontology quality assessment and ontology quality optimization, to be validated on existing and new ontologies in the cybersecurity domain.

1.4. Thesis overview

This thesis is organized into five chapters, namely Introduction, Literature Review, Methodology, Analysis and Conclusions.

In the first chapter, the motivation and objectives of the developed work is presented.

Chapter two presents the theoretical background of the ontologies used in terms of state-of-the-art technology and also with the methodologies and techniques that gave support for the conclusion of this thesis. This chapter includes a thorough explanation of the ontology assumptions of the study, a taxonomy of different cyberattack vectors is presented. Next, the usefulness ontologies are presented by two examples. Furthermore, a description is given of the process.

Chapter three described the research methodology in this study. Chapter four described in detail the most relevant ontology quality metrics found in this study, it synthesizes the ontology of quality improvement and presents the resulting ontology for vulnerability characterization and its implementation.

The thesis concludes with an outlook on further developments of OWL and points out future directions on the methodological challenges of developing the ontology and on the potential application for the incident response team, in particular by involving the research and practice community.

2. Literature Review

2.1 Ontology

2.1.1. Ontology concept

A major landmark in the study of ontology was introduced by Neches, Fikes, Finin, Gruber, Senator & Swartout. These authors clarified that an ontology defines the terms and relationships comprising the vocabulary of a topic in an area. Ontology also defines the rules for combining terms and relations to define extensions of the vocabulary [3].

According to Gruber, ontology is an explicit specification of a conceptualization, meaning an idea that a person or a group can have of the world [4]. Ontologies specify the components of concepts, relations, functions, instances, and axioms. Gruber added that ontology is a formal, computer-readable specification in which the elements are shared concepts and represent a consensual and accepted knowledge by a group of people [5].

In addition, this includes definitions and an indication of how concepts are interrelated. Collectively they impose a structure on the domain and constrain the possible interpretation. Ontology was defined as a science of being, of the types and structures of objects, properties, events, processes and relations in each area [6].

For a better understanding of cyberspace and its concepts, Rees shows that ontology constitutes concepts that describe a particular domain. Such concepts are defined through a hierarchy of subclasses with attributes and relations between concepts [7].

In computer science, ontologies are methods of knowledge representation, having been developed in the specific field of Artificial Intelligence to facilitate the sharing and reuse of knowledge [8].

Evidence reveals that the systems have their own implicit ontology. This ontology assigns meaning to the symbols used according to a particular world view. However, an ontology may have different roles in a system [2]. The availability of knowledge stored in ontologies can provide the mechanisms required to organize, store, and access information for items that include database schemas, user interface objects, and application programs.

The use of ontologies in information systems allows the establishment of correspondence and relations between the different domains of information entities.

Otherwise, systems are essentially artifacts of knowledge that capture and represent knowledge about certain domains.

While ontologies are used to specify and communicate the knowledge of a domain, there is a recognition of the implicit ontological principles and concepts when they are assigned meaning to the symbols used and can have different roles in an information system [2].

Studies have shown that the word ontology has several definitions. An ontology provides notions of the basic concepts of a domain, appropriate for automatic processing. For the semantic web, ontologies (described in Ontology Web Language - OWL) are very useful, consisting of structures with classes to represent the general concepts of any area, relationships identified between the objects, properties or attributes of the objects described. Ontologies facilitate the research of information and data integration of different communities. This is because they provide a common basis for ensuring data consistency, which are properly categorized in a standardized way. Ontologies can be applied to improve the functioning of existing web applications and allow the implementation of new applications and services. We can verify the use of ontologies in an environment of great interoperability, with the normalization of concepts and automatic processes for consultation and information exchange [9].

2.1.2. Ontology classifications

Borst pointed out that ontology is a formal specification of a shared set-up, in which an ontology should be read by a machine and must reflect the consensual knowledge accepted by a group [10].

Ontologies can be classified into two formats. In a horizontal ontology we try to reach a representation of all possible concepts, without a very detailed description. In the case of a vertical ontology, there are only concepts of a specific area, with a complete description, according to the domain in which they are inserted.

On the other hand, Van Heist [11] proposed another classification of ontology. This terminology specifies the representation of knowledge in the universe of discourse that is used to unify vocabularies in a given domain. The information ontologies detail the structure of database storage and provide a framework for this. In addition, knowledge modelling ontologies highlight the cognition of knowledge. These also contain a solid internal structure that can be tailored to the particular use it describes.

Evidently, this view allows distinguishing ontologies based on the degree of formality in the specification. Informal ontologies use a natural language. Semiformal ontologies provide axioms such as taxonomies. Formal ontologies define vocabulary semantics by a complete and effective axiomatization.

According to the level of dependence of a task or vision, Mizoguchi, Vanwelkenhuysen, & Ikeda [12] present several classifications of ontologies. High level ontologies describe general concepts and are independent on a particular problem or domain. Domain and task ontologies describe the vocabulary related to a generic domain or a specific task or activity, with a specialization of terms introduced in the high level ontology. Application ontologies describe concepts that depend on both a domain and a particular task. These concepts correspond to functions performed by domain entities in performing certain activities. Such concepts contain knowledge essential to modeling a specific application.

Also, the ontology can have a variety of forms, a vocabulary of terms and some specifications of their meaning [13]. Through the web it has reached a stage of collective intelligence where everyone has the possibility to share information [14].

2.1.3. Knowledge representation ontologies

In the literature, at least five different types of knowledge are distinguished and the knowledge model components are related to task goals [11].

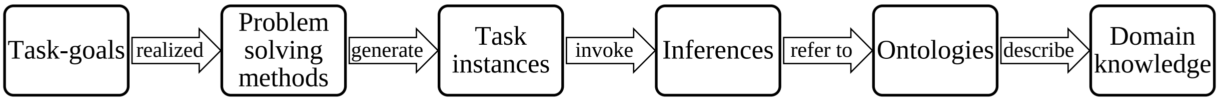


Figure 1 – Schema of the knowledge [11]

According to Sheth, in the knowledge gaps it is necessary to solve semantic interoperability. What is accomplished can be more than a description in order to make good use of the information available on the internet and in distributed computing [15].

The World Wide Web Consortium (W3C) has developed common protocols and guidelines to ensure web interoperability, to promote the participation of anyone, to share knowledge and to create a global trust base. Informal languages such as natural language are expressive, but they generate ambiguous interpretations. Formal languages provide the creation of models with reduced ambiguity and with consistent meanings for the context of the organization [16].

Ontologies are methods of knowledge representation, having been developed in the field of Artificial Intelligence to facilitate the sharing and reuse of knowledge [8]. Besides, ontologies are known for being a good way to formalize knowledge. It is well known that ontologies are useful for representing and inter-relating many types of knowledge of a domain [17].

Figure 2 shows an overview of the OWL 2 standard.

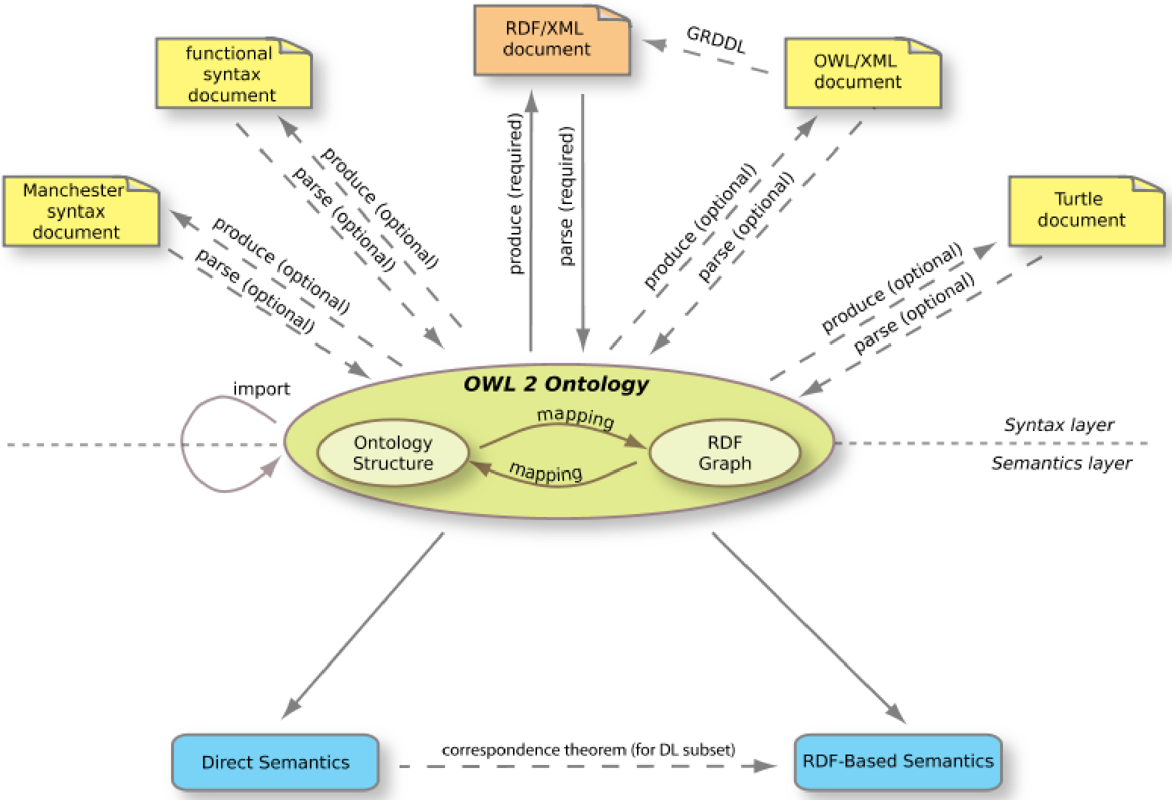


Figure 2 – An overview of the OWL 2[1]

2.2. Semantic web

The goal of the semantic web is the development of standards and technologies that allow machines and automatic processes to understand more information expressed on the web. This serves as a support for knowledge discovery, task automation and data integration. These data have a correspondence with reality that can be established through a semantic for the information technologies [18].

[1] Figure from the website: https://www.w3.org/TR/owl2-overview/

In this context, as shown in figure 3 a semantic is a logical structure that links the representation of an object in a database and its object in the real world [19].

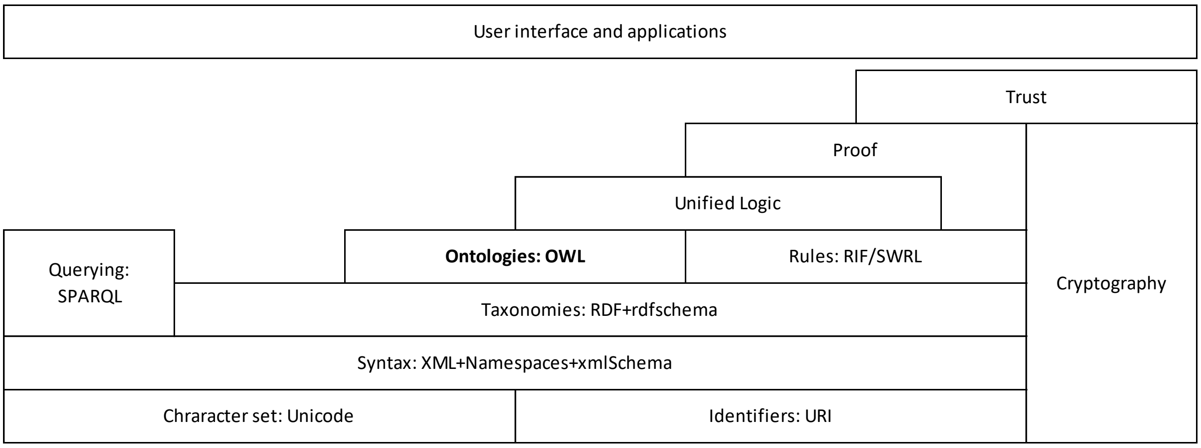


Figure 3 - Diagram of the layers of the semantic web architecture

The Unicode and URI standards layer is related to the definition of appropriate characters and object referencing mechanisms by their address. At the level of the XML layer there is a description based on XML standards, namespaces and xml schema that makes it possible to integrate semantic web definitions into other applications. The top layers: logical proof, confidence and cross-way digital signatures are presented only as a demonstration. In the trust layer there are mechanisms to trust or not based on certain evidence. Finally, there are two more intermediate layers, RDF, RDF Schema and ontologies. The first one encompasses semantic web languages, defining the vocabulary used to describe the objects and their types. Then there is a layer of hierarchical structures of objects, usually grouped by classes, defined with a language of the previous layer. The inclusion of a document to the semantic web aims only to add semantic annotations to the initial format, in order to facilitate the automatic processing of the same information. This format is perceived by machines and must have rules and guarantees to be used correctly in a distributed database [20].

In addition, the web has gained in size from the semantic web and the web of data, with the objective that the computers meet relevant outcomes. The developed systems can support reliable interactions in the network and the data can provide links to other resources [21, 22].

While semantics are related to the understanding of the nature of meaning, the semantic web by its standards must ensure that the contents are intelligible both by the human entity and by the machine entity [23].

As the semantic web grew, several languages emerged to encode the semantics of documents. The main languages that marked the evolution of the semantic web were: Resource Description Framework (RDF), RDF Schema (RDFS), Topic Maps which resulted in XML Topic Maps (XTM), Simple HTML Ontology (SHOE), Ontology Inference Layer (OIL), DARPA Agent Markup Language (DAML), DAML & OIL and the latest Web Ontology Language (OWL), which is a semantic markup language for publishing information on the Web, used to represent concepts and relationships between concepts. The OWL language is an extension of RDF and derives from the DAML and OIL language, incorporating the updates resulting from the design and application experience of that language [9].

2.3. Ontology web language

Among the various languages for describing ontologies, the most widespread and most relevant is OWL, which has been improved and extended through OWL 2 [9]. The OWL language was defined by W3C in three sublanguages that allow users to choose the most appropriate balance between expressiveness and support for reasoning [9]. For W3C, each sublanguage is an extension of the previous one and in each application the most suitable one is chosen, according to the expressive parameters, as follows:

- OWL Lite. For the basic needs of a hierarchy and simple constraints, in addition to applying the constraints present in the other variants that guarantee computational efficiency and support for reasoning.
- OWL Description Logics - DL. When it is desired to reach the maximum expressiveness of a description language that recovers the support to the reasoning and computational efficiency lost in OWL Full. This strand uses some restrictions as limitations to the use of RDF Schema. In other words, it restricts OWL primitives to RDFs and requires type disjunction.
- OWL Full. When one wants the maximum expressiveness and the syntactic freedom of the RDF, without computational concerns. OWL Full is the most complete of the three sublanguages because it comprises the entire OWL, in which the use of all its primitives is valid and unrestricted. Its advantage is that it is fully compatible with RDF Schema without imposing restrictions on an RDF or RDF Schema document so they can be considered as OWL Full documents. In this way, the user is responsible for solving any problem arising from the execution of the reasoning [9, 24].

An ontology provides notions of basic concepts of a domain, appropriate for automatic processing. For the semantic web, ontologies are very useful, consisting of structures with classes to represent general concepts of any area, relationships identified between the objects, properties or attributes of the objects described. Ontologies can be applied to improve the functioning of existing web applications and allow the implementation of new applications and services. We can verify the use of ontologies in an environment of great interoperability, with the normalization of concepts and automatic processes for consultation and information exchange [9, 25].

2.4. Cyberspace

Cyberspace is a complex environment resulting from the interaction of people, software and services on the Internet by means of electronic devices and networks connected to it. (ISO/IEC 27032:2012).

Cyberspace provides nowadays the means for essential human activities such as social dynamics, economics, industrial and cultural activities, etc. Many of the threats present in the physical world moved to the cyberspace and present similar or more severe harmful effects. In addition, due to the nature of cyberspace, new non threats have emerged that need to be clearly defined, understood by professionals and practitioners in the cybersecurity area and, whoever possible, dealt with in an automatic way by computational means.

Because in this thesis we are interested in studying ontology-based knowledge representation and its application in the cybersecurity domain, in the next section we present core concepts, relations, rules/axioms in this domain, followed by the corresponding OWL ontology representation in further sections.

2.4.1. Cybersecurity

Security is a collection of technical approaches that address issues covering physical, electronic and procedural protection for information. Security should include identification of potential threats to systems and data. This encompasses protection of data from inadvertent or malicious inappropriate disclosure and the non-availability of data due to system failure and user errors.

Security Measurements are used to facilitate decision making and improve performance and accountability through the collection, analysis and reporting of relevant cybersecurity performance-related data. The purpose of measuring performance is to measure activities in order to improve them by applying corrective actions based on observed measurements (NIST SP800-55).

Network security is concerned with hardware, software, basic communication protocols, network frame structure, and communication mechanisms factors of the network. Information Security in the network context deals with data integrity, confidentiality, availability and non-repudiation while sent across the network (ISO/IEC TR 29181-5).

Cybersecurity comprises resources, processes and controls that are designed to protect networks, systems and data from attack vectors. Cybersecurity is a holistic area aiming to ensure confidentiality, integrity and availability of information in cyberspace (ISO/IEC 27032:2012).

Cybercrime is a criminal activity where services or applications in cyberspace are used for or are the target of a crime or where the cyberspace is the source, tool, target or place of a crime (ISO/IEC 27032:2012).

The cyber threat landscape continues to evolve, becoming more technically complex. Advanced attack strategies and tools are developed with increasing sophistication, such as malware, armed with zero-day exploits, that autonomously targets vulnerable devices and spreads with little human intervention, likely to overpower an already challenged information or operation security. Table 1 presents a summary of core concepts related to cyber-attacks.

Table 1 – Concepts of cyber-attack vectors

Term	**Description**
Attack	An attack is an intrusion that violates the security policy of a system. It attempts to destroy, expose, alter, disable, steal or gain unauthorized access to or make unauthorized use of an asset (ISO/IEC 27000:2016).
Vulnerability	Weakness in the security system, in procedures, design, or implementation etc. that might be exploited to cause loss or harm.
Threat	Potential for an event that can breach security and cause harm.
Asset	Target of the threat that potentially results in a loss of value.
Risk	Expectation of loss expressed as a probability that a particular threat will exploit a certain vulnerability that will result in a harmful result [26].
Cross-site scripting	The most well-known type of vulnerability found in web systems. In this technique the attacker puts malicious script into the web to access the main server [27].
Denial of Service (DoS)	Prevents or inhibits the normal use or management of communications facilities.
Distributed Denial-of-Service (DDoS)	An attempt to make a machine or network resource unavailable for its intended use. It often consumes more computer resources than a device can handle or disrupts by disabling communication services.
Malware	Malicious software designed specifically to damage or disrupt a system, attacking confidentiality, integrity and/or availability (ISO/IEC 27033-1:2015).
Man-in-the-Middle	When a perpetrator positions himself in a conversation between a user and an application—either to eavesdrop or to impersonate one of the parties.
Phishing	An attempt to acquire sensitive information or to make somebody act in a desired way by masquerading as a trustworthy entity in an electronic communication medium.
Ransomware	Malicious software that demands a ransom fee be paid after the software is installed on a computer system.
Spoofing	Providing false information about an identity in order to gain unauthorized access to systems.
Social Engineering	Process of exploiting the weakest link, the people, in the system with illegitimate motivations.
Trojans	Program that appears to be a certain function, but is actually performing malicious activity when executed.
Virus	Malicious code that is loaded onto a computer without the user’s knowledge. It can replicate and spread to other computers by attaching itself to another computer file.

2.4.2. Cybersecurity ontology

Syed, Padia, Mathews, Finin & Joshi describe ontologies that provide a common understanding of a cybersecurity domain and unifies most commonly used cybersecurity standards [28]. The cybersecurity ontologies can act as first filters allowing for contextual understanding of cyberspace activities, events, stimulus and signals, including concepts that transcend network and information security, namely, notions about people, time, space, etc.

The ontology should represent which threats endanger which assets and what countermeasures can lower the probability of the occurrence of an attack [26].

In cyberspace there are actions that can be performed by anyone with a minimum of computer knowledge in order to intentionally or unintentionally harm any other individual or organization. Using cybersecurity policies, it is possible to understand the concepts and security needs of cyberspace. Such policies are used in the context of the security and continuity plans of the activity of organizations. State-of-the-art cybersecurity results from the use of information and communication technologies, transparency measures, cooperation measures and stability measures to improve cybersecurity response capability [29].

Figure 4 shows a simplified version of an ontology proposal that covers aspects that were not considered in prior works and provides an improved knowledge representation of this domain.

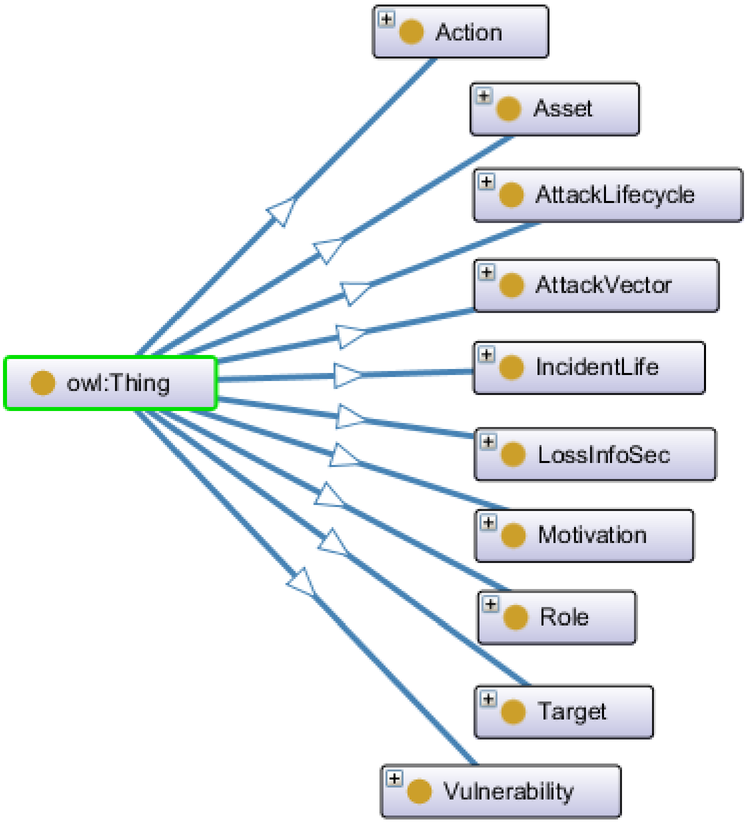


Figure 4 – Ontology in cybersecurity with Protégé editor version 5.5.0-beta-3

2.5. OWL ontology in the cybersecurity domain

Although several areas in the cybersecurity domain are covered by OWL ontologies, for the sake of simplicity and clarity, we concentrate on ontologies focusing on cyberattacks. The ontology research literature offers a number of attempts to create taxonomies and conceptual models of cyberattacks and attack patterns.

The Common Attack Pattern Enumeration and Classification (CAPEC^{TM}) provides a publicly available catalogue of attack patterns along with a comprehensive schema and classification taxonomy. The mechanisms of attack represent patterns in a hierarchy that are employed to exploit a vulnerability [30].

The main threat faced by information systems and their users is malicious software referred to as malware. Malware samples became very complex pieces of code that leverage a broad range of techniques to attack computer systems. These attacks aim to compromise systems. Malware can install itself, establish remote access by its controller, bypass the security mechanisms, and finally, accomplish its objective. Malware is translated into instructions that can be seen as an action performed over a specific resource of the infected system. The resulting set of actions corresponds to the "*behavior*" of a malicious program. Knowledge about malware behavior is the key to planning more secure systems and preventing future attacks. Obtaining and analyzing behavior associated with malware is one effective way of understanding infection procedures.

Silva & Rodriguez listed 41 general security ontologies [31]. In addition to this list, 52 more have been identified [32]. Accordingly, a brief description of the reported ontologies has been presented in the table 2 along with the related discussion as analysis.

Table 2 – Cybersecurity ontologies [31, 32]

Nº	Cyber Ontologies	Cybersecurity Domain	Referencies
1	DAML+OIL and DAMJEssKB	Computer attacks for sharing knowledge of intrusion detection	[33]
2	Ontology for computer network attacks	Distributed IDS	[34]
3	A ontological structure for information security	Information security standards, security policies and control	[31]
4	OWL-based ontology of information security	Information security	[35]
5	Swimmer's malware class hierarchy	Malware	[30]
6	An integration of ontology-based and policy-based approaches	Automate pervasive network security management	[31]
7	Common Weakness Enumeration (CWE)	Software Security weaknesses	Mitre
8	Representation based on the knowledge of ontology	Distributed multi-agent peer-to-peer IDS	[31]
9	An intelligent system with an ontological bases that analyzes the input	Detect attacks	[31]
10	An ontological approach	Concepts of information security	[31]
11	A modeling ontology for integrating vulnerabilities	Vulnerabilities Security requirements	[36]
12	Semantic technology to information security	Software vulnerability management	[31]
13	Security ontology OWL+SWRL+OWL-S	Knowledge and information Contextual alert analysis	[31]
14	An ontology based approach	Security policies	[37]
15	A security ontology for incident analysis	Incident	[38]
16	Hierarchical model of alert correlation knowledge and the XSWRL ontology technique.	Intrusion alert correlation system	[31]
17	A Security Audit Framework	Information security	[39]
18	Common Vulnerability Scoring System (CVSS)	Security metrics	[26]

19	Network Intrusion Prevention System based on Ontological and Slow	Intelligence	[31]
20	Ontologies for Security Requirements	Security requirements	[40]
21	Security ontology for concepts	Attacks, countermeasures, security properties	[31]
22	An ontology-based attack model	Attack impact, attack vector, attack target, vulnerability and defense	[41]
23	Security ontologies with OntoMetric	Security standards	[31]
24	A Security ontology with MDA	Security, Software development	[42]
25	Concepts of the semantic web and ontologies for analyzing security logs	IDS software, Index of false positives and false negatives	[31]
26	Cyber-security ontology using textbook index terms	Security textbook	[43]
27	A Packet-Centric Network Ontology of Cyber Defense	Network traffic	[44]
28	An ontological approach	Vulnerabilities and attacks	[31]
29	An ontology-based problem-solving system for cyber-attack management	Identify and defend against cyber attacks	[45]
30	An ontological engineering methodology	Design and evaluate security systems	[46]
31	Fusion model comprised of class keys	Network environment, vulnerability, attack, security incident and sensors	[31]
32	IDS ontological model	Types of attacks and vulnerabilities	[31]
33	An ontological representation of a network	Specification-based IDS	[47]
34	A web-based tool for network management control of network	Data sources Artificial intelligence	[48]
35	An ontology from a database of cyber security knowledge graphs	Structured and unstructured data sources	[49]
36	MAECTM: Malware Attribute Enumeration and Characterization	Malware (Mechanisms and behavior)	Mitre

37	Common Attack Pattern Enumeration and Classification (CAPEC)	Data on attacks Attack patterns	Mitre
38	Cyber Observable eXpression Archive Website	Cyber observables	Mitre
39	Modeling Enterprise Level Security Metrics	Security and Threats	[26]
40	Integrated Cyber Analysis System (ICAS) ontology	Incident response	[50]
41	OASIS Structured Threat information eXpression (STIX)	Cyber threat intelligence	Mitre
42	CMU Insider Threat Indicator ontology	Threat indicators	[51]
43	Comprehensive ontology in network security	Network security	[31]
44	OWL ontology for cybersecurity	Corporation Security requirements	[52]

Martín (2018) created the following OWL cybersecurity ontologies including the following classes and axioms (Figure 5) [53].

$DoSAttacksNetwork \equiv DisallowedNetwork \sqcap \geq 2000connectionsPerSecond$ (1)

$HotAttacksNumberNetwork \equiv DisallowedNetwork \sqcap \exists phoneNumberLiteral$(2)

$DisallowedNetwork \ \equiv Network$ (3)

$MultipleFailedLoginUser \ \equiv DisallowedUser \sqcap \geq 4failedLoginAttempts$ (4)

$BlockedUser \ \equiv DisallowedUser \sqcap \exists identifier.Literal$ (5)

$DisallowedUser \equiv User$ (6)

$DisallowPolicy \ \equiv DisallowedNetwork \ \sqcup DisallowedUser$ (7)

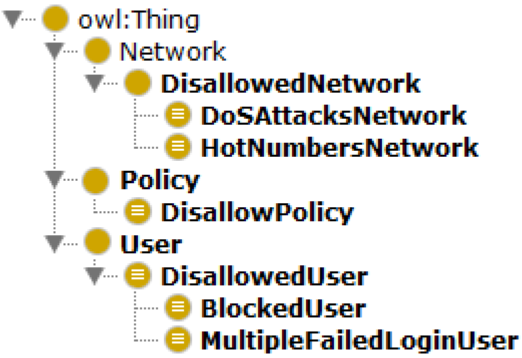


Figure 5 – Classes of an ontology [53]

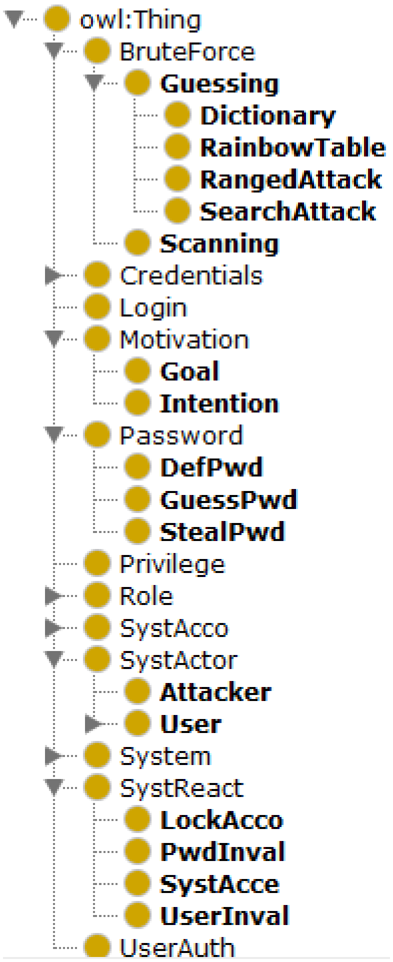
A brute-force attack is an excessive number of authentication attempts per event by the same or different user(s). These may be exploited to compromise credentials such as usernames or passwords by brute force [54]. The brute-force attack is often the last step in password cracking. It needs an attacker to try all possible combinations of characters within the length of the password. A short 4-letter password consisting of lower-case letters can be cracked in just a few minutes. However, a 7-character long password consisting of either upper or lower case letters would take 267 guesses. A combination of alpha-numerical characters along with case-sensitivity and special characters would increase the complexity significantly. This might be impossible to crack within a reasonable period of time. For Bojanova, Yesha & Wu (2016), these attacks are represented using a combined model in the following way where PassCrack denotes the generic password cracking attack:

$$PassCrack \triangleq < a\,\{D\} \rightsquigarrow u, HYB, \{AL\}, \{medium, high\}, ACTIVE > \quad (8)$$

A guessing attack is the simplest approach for password cracking to guess the password. Many unaware users do not comprehend the need to maintain a secure difficult-to-guess password and hence, they often use passwords which can be easily guessed. A few examples of such easily guessable passwords are to use the word "password", the same password as the username or easily identifiable names as the passwords. This approach is more effective if the attacker personally knows the victim and has knowledge of such information which is susceptible to password guessing [55].

Another method is the dictionary attack. In this attack, the attacker utilizes a program or a script that tries different possible combinations of words in a dictionary along with some additional special characters (such as '\#', '\$', '_' and so on) often used in the beginning or in the end of a password [55]. The attacker usually possesses a copy of the English dictionary as well as foreign language dictionaries for this purpose. In addition, dictionary-like databases containing names and lists of common passwords are often used.

Figure 6 presents the classes and sub-classes that identified the components of attack scenario. The structure focus on brute force attack, to provide a better level of specific scenario.



- owl:Thing
  - BruteForce
    - **Guessing**
      - **Dictionary**
      - **RainbowTable**
      - **RangedAttack**
      - **SearchAttack**
    - **Scanning**
  - Credentials
  - Login
  - Motivation
    - **Goal**
    - **Intention**
  - Password
    - **DefPwd**
    - **GuessPwd**
    - **StealPwd**
  - Privilege
  - Role
  - SystAcco
  - SystActor
    - **Attacker**
    - **User**
  - System
  - SystReact
    - **LockAcco**
    - **PwdInval**
    - **SystAcce**
    - **UserInval**
  - UserAuth

Figure 6 – Classes of a cybersecurity domain

There are many reasons behind the discrepancy between demand and supply of semantic models of cybersecurity. A great part of the problem is the lack of balance between the vertical and the horizontal directions of the ontology structure. The structure typically yields rich catalogs of cyber-attacks, exploits and vulnerabilities. On the other side, a rigorous conceptual analysis of the entities and relationships that are encompassed by different cyber scenarios would also be needed to explore in depth the semantic area of operations [56].

This thesis proposes the identification of metrics to assess the quality of ontologies. It has a particular interest in quality assessment of currently available ontologies in the cybersecurity domain.

In the chapter four a set of ontology quality assessment metrics are presented. These metrics play an essential role in the decision process of selecting ontologies for different purposes.

When someone needs to select and use an ontology for learning/teaching purposes or to integrate an ontology as part of a component of a software (expert) system, he or she faces a difficult decision/selection problem. The amount of available ontologies in all domains, including in the cybersecurity domain, requires strategies, techniques and tools to assist the ontologies selection process.

The following chapter presents the methodology of study and the chapter four presents a set of reference quality assessment ontology metrics that can be of help in the selection process.

3. Methodology

The aim of developing a quality assessment proposal, taking as a detailed example an ontology in the cybersecurity field, is to put in context ontologies quality assessment research and, complementarily, conceive a solution-oriented structure, through the definition of semantic rules, to allow for a process of comparability across different vulnerability and applied concepts and methods.

The development of the ontology and the implementation of the semantic was realized in a three step iterative process that is built progressively upon each step. First, a research of existing literature, data, models, and methods to conduct quantitative and qualitative assessments of cybersecurity vulnerability was performed. In this way, we obtained an overview of the use of vulnerability in different research fields and extracted relevant classes and categories to structure the semantic fields. The second step was developing the ontology itself, which allows for the explicit description of methods, concepts, and models that are useful for the classification of vulnerability assessments.

This research work follows the Design Science Research Methodology (DSRM)[57]. DSRM defines the following cycle of steps:

- Identification of the problem and motivation - Defining the context of the specific research problem and justifying the value of the solution through an exploratory-descriptive study.
- Definition of the objectives towards the solution - Infer the quantitative and / or qualitative objectives of the solution by defining the problem and the work performed.
- Design, development and implementation of the solution - Develop the artifact with the desired functionalities and incorporate in its design the contribution of the research.
- Demonstration - Demonstrate the artifact to solve instances of the problem.
- Validation and evaluation - Compare the objectives of a solution with the results observed and obtained through demonstration of the artifact.

The techniques studied in this thesis are fundamentally based on the analysis of characteristics such as the structure and content of OWL ontologies, through the theory of graphs, predicates logic and optimization techniques.

The standards and support tools for the construction of ontologies will be standard: W3C OWL, OWL API and Protégé software. To validate this research, ontology quality metrics and ontology quality improvement formulations will be presented with highlighted applications in the area of cybersecurity.

This aims to contribute to the improvement of the results of the ontology research initiatives in this domain, using a standard ontology representation language, to support organizations in their operations and strategic cybersecurity decision making processes.

4. Analysis

4.1. Ontology quality metrics

The emergence of ontologies as a standardized way of representing knowledge created an increasing interest in the subject and resulted in the emergence of a high amount of ontologies across all areas of knowledge, including in the cybersecurity domain. These address a wide range of topics, such as vulnerabilities, exploitation, malware and incidents.

Recent developments in natural language processing, ontology design and engineering also made available tools for the automatic generation of ontologies. These are automatically created from natural language texts/documents [59]. This trend of increasing the ontologies manual and/or automatic creation leads to the formation of defects in ontology structure and content. The complex structure of the relationships between concepts represented in the ontology requires the application of quality assessment metrics, procedures and practices for quality control. Conflicts and quality issues in ontologies may arise due to the data/knowledge coming from a variety of sources. Other quality requirements and criteria related to the intention of use of the ontology must also be taken into account. For instance, adapting an ontology content to its intended use by excluding resources that are rarely or not used, as well as resources not belonging to a particular subject area of interest.

In the remainder of this chapter, a set of ontology quality assessment metrics are identified, described. Metrics with a quantitative nature will have their computation details formalized and explained.

The variables and metrics definition are presented below, follow the SquaRE ontology assessment [60]:

- Annotation of class C_i: A_{C_i}
- Classes of ontology: C1, C2, …, Cn
- Instances of class C_i: I_{C_i}
- Properties of class $C_{i:}$ P_{C_1}; P_{C_2}; …; P_{C_n}
- Restriction of class C_i: Att_{C_i}
- Relationships between classes: R_{C_1}; R_{C_2}; …; R_{C_n}
- Super classes directions of class $C_{i:}$ Sup_1; Sup_2; …; Sup_n

Annotation richness (ANOnto). Number of annotation properties per class.

$$ANOnto = \sum|A_{C_i}|/\sum|C_i| \tag{9}$$

Attribute richness (AROnto). Number of restrictions of the ontology divided by the number of classes.

$$AROnto = \sum|Att_{C_i}|/\sum|C_i| \tag{10}$$

Class richness (CROnto). Number of individuals per class.

$$CROnto = \sum|I_{C_i}|/\sum|C_i| \tag{11}$$

Coupling between objects (CBOOnto). An average of direct parents per class minus the relationships of the Thing class.

$$CBOOnto = \sum|Sup_{C_i}|/(\sum|C_i| - |R_{Thing}|) \tag{12}$$

Depth of Inheritance Tree (DITOnto). Length of the largest path from Thing to a leaf class of the ontology.

$$DITOnto = Max\ (\sum D|C_i\ |) \tag{13}$$

Lack of Cohesion in Methods (LCOMOnto). The length of the path from the leaf class to Thing, divided by the total number of paths in the ontology.

$$LCOMOnto = \sum path[|[C\ (Lea\ f)_i|]]/m \tag{14}$$

Number of children (NOCOnto). Mean number of the direct superclasses per class minus the subclasses of Thing.

$$NOCOnto = \sum|R_{C_i}|\ /(\sum|C_i|\ -\ |Sup_{C(Lea\ f)_i}|) \tag{15}$$

Number of properties (NOMOnto). Mean number of datatypes properties and object properties per class.

$$NOMOnto = \sum |P_{C_i}| \sum |C_i| \qquad (16)$$

Relation Richness (RROnto). Number of the properties in each class are used at the instances level.

$$RROnto = \sum|P_{C_i}|/\sum(|R_{C_i}| + |P_{C_i}|) \qquad (17)$$

Relationships per class (INROnto). Number of subclasses per class.

$$IROnto = \sum|P_{C_i}|/\sum(|R_{C_i}| + |P_{C_i}|) \qquad (18)$$

Response for a class (RFCOnto). Number of datatype properties and object properties that can be directly accessed from the class.

$$RFCOnto = (\sum(|P_{C_i}| + \sum|Sup_{C_i}|)/(\sum|C_i|) \qquad (19)$$

Considering that $\boldsymbol{C(MDP)_i}$ is a number of classes with more than one direct father:

- **Tangledness (TMOnto).** An average of class with multiple parents.

$$TMOnto = \sum|C(MDP)_i| \,/\, (\sum|C_i|) - 1) \qquad (20)$$

- **Tangledness 2 (TMOnto2).** An average of parents directly by class, of classes that have more than one direct parent.

$$TMOnto2 = \sum|C(MDP)_i|/\sum|C(MDP)_i| \qquad (21)$$

Direct Parent average (DPOnto). An average of direct parents by class.

$$DPOnto = \sum|C(MDP)_i|/\sum|C_i| \qquad (22)$$

Weighted Method Count (WMCOnto). Number of datatype properties, object properties and subclasses per class.

$$WMCOnto = \sum path[C(Lea\,f)_i\,|]/\sum|C(Lea\,f)_i| \qquad (23)$$

Compatibility. The ability of two or more ontologies to exchange information and/or to perform their required functions while sharing the same environment.

Consistency. Concepts and relationships must have durability and meet requisite ontological standards.

Empower. One ontology to refer explicitly to another. For example, triples from the imported ontology are available for inference.

Expressivity. The ability of a modeling language to describe certain aspects. More expressive modeling language can express a wider variety of statements about the model. All modeling languages of semantic web differ in their levels of expressivity [23].

Maintenance. The capability of ontologies to be modified for changes in environments, in requirements or in functional specifications.

Operability. Strength needed for use, and in the individual assessment of such use, by a stated or implied set of users.

Performance. The level of performance and the relationship with the resources used, conditions and it includes data load time, query response time, memory consumption and scalability [61].

Reliability. Capability of ontologies to maintain their level of performance under stated conditions for a given period of time.

Reusability. The access to libraries of reusable ontological components would facilitate the knowledge engineering process [11]. The underlying idea is that some concepts are more reusable than others, and that the reusability of concepts depends on how specific these are for particular domains and how specific these are for particular problem-solving methods [11] and many knowledge-representation systems can import and export ontologies.[62].

Sharing. To enable the sharing of ontologies, they must be explicitly described in a way understandable to all relevant agents.

Similarity. This refers to the similarity between components of the entity descriptions and similarity between graphs representing entities. A map will be created for two entities in two different ontologies when these entities are considered semantically similar, or, when their similar value is higher than a certain threshold. [23].

Transitivity. "Transformation rule for transitivity axiom is also expressed regarding object's attributes" (Hnatkowska, 2018: 9) [63]. The ontology can be transferred from one environment to another.

Temporal representation. The ontology provides a common understanding of a domain and uses a very basic representation of time where time is represented as a data property associated with classes that represent events so that can support temporal reasoning. [28].

Variability. When describing a set of things, some of them will have some things in common, and some will have important differences. Managing variability is a fundamental aspect of modeling in general and of semantic web models.

Often a design pattern consists of only a single statement but one that is especially helpful when used in a particular context. The value of the pattern is not so much in the complexity of its realization but in the awareness of the sort of situation in which it can be used. The viewpoints will be overlaps, disagreements and confusion before there is synergy, cooperation and collaboration in a web of knowledge. On the semantic web we don't know in advance how information from somewhere else on the web should be interpreted in a new context. In OWL, it is possible for the class structure to change as more information is learned about classes or individuals [23].

Figure 7 illustrates[2] different classes that the brute force attack can perform in the action and reaction components. The spherical representation about threats of the system uses vulnerabilities like social engineering, dictionary and brute force attacks which exploits some vulnerabilities. According to the suggested metrics the dynamic of security level of the graph is evaluated.

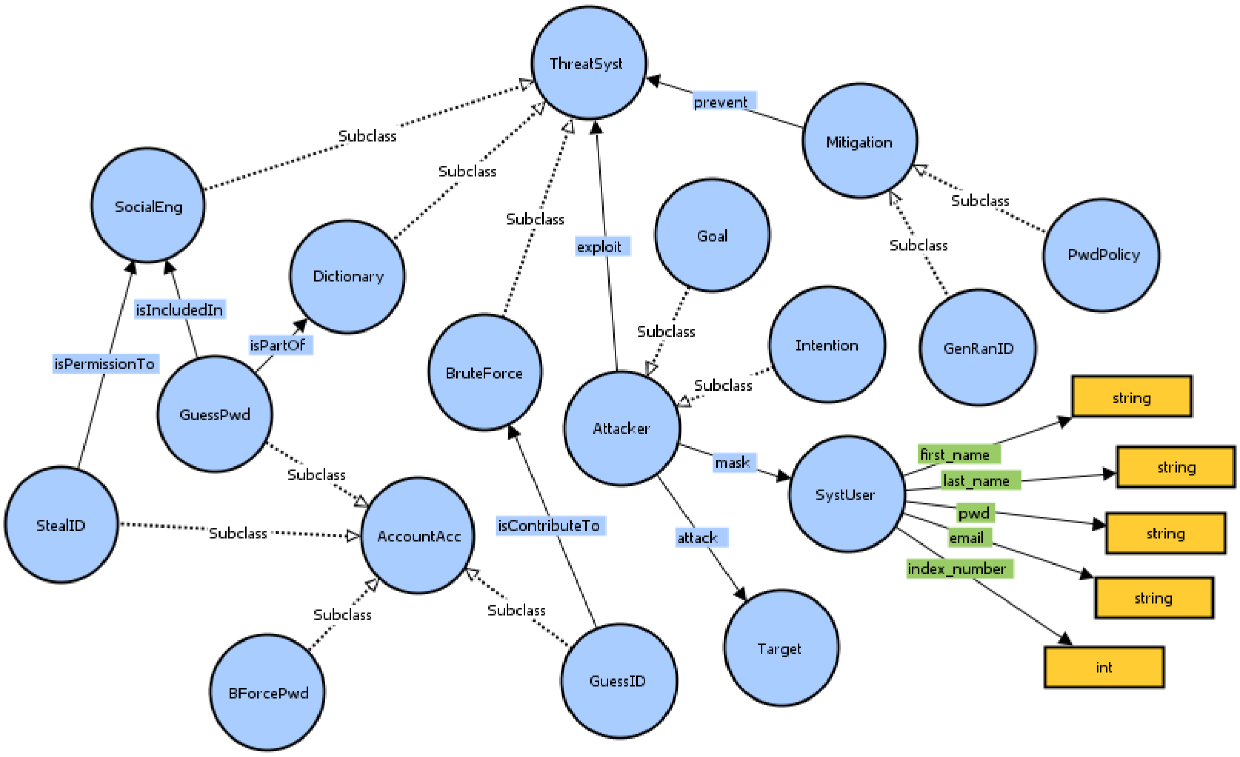


Figure 7 – Results from the VOWL plugin of an OWL ontology structure #1

[2] Ontologies design provided by Protégé OWL ontology editor, version 5.5.0-beta3 and plugin VOWL, version Beta 0.1.4.49.

In Table 3 is possible to observe the domain and the range associated with object properties

Table 3 – The object properties of the ontology #1

Domain	**Object Properties**	**Range**
Attacker	attack	Target
Attacker	exploit	ThreatSystem
GuessID BForcePwd	isContributeTo	BruteForce
BForcePwd GuessPwd	isIncludedIn	SocialEng
GuessID GuessPwd	isPartOf	Dictionary
StealID BForcePwd	isPermissionTo	SocialEng
Attacker	mask	SystUser
Mitigation	prevent	ThreatSystem

Table 4 contains the domain and the range in each datatype properties defined of the ontology #1.

Table 4 – The datatype properties of the ontology #1

Domain	**Datatype Properties**	**Range**
SystUser	email	String
SystUser	first_name	String
SystUser	index_number	Integer
SystUser	last_name	String
SystUser	pwd	String

The generated OWL ontology is shown in Figure 8. The first observation is the vertical structure with the main classes. Other interesting aspect from the ontology #1 is the difference of positions. As can be verified in Appendix 1 the number of classes, object and data properties show more persistence and capability in actions and reaction operations. This ontology expresses a wide variety of combinations of classes and situations.

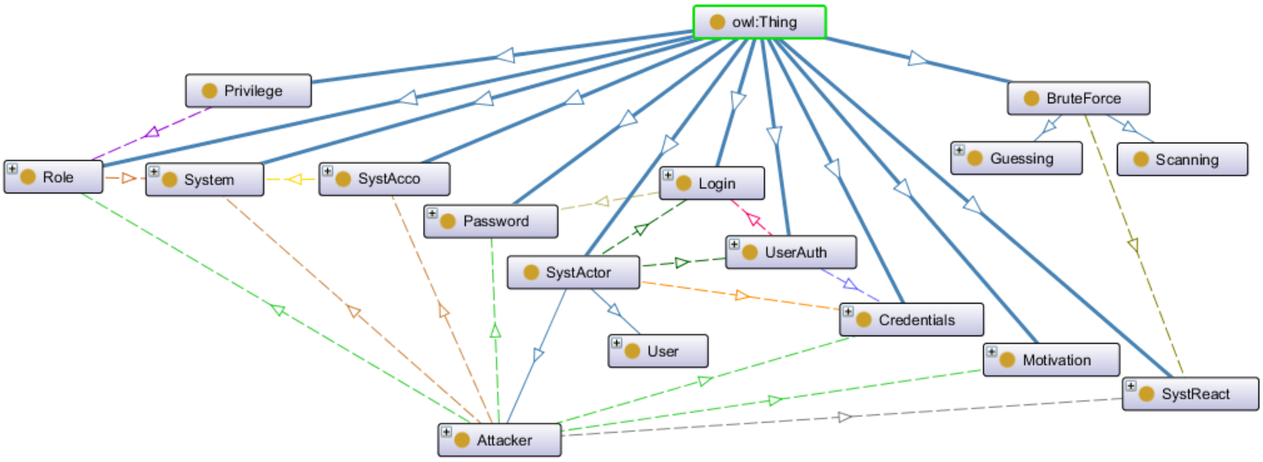


Figure 8 – OWL ontology structure #2

Tables 5 and 6 exhibit the object properties and datatype properties of the generated ontology respectively.

Table 5 – The object properties of the ontology #2

Domain	**Object Properties**	**Range**
Attacker	attacks	Victim
TrueID	contributes	UserAuth
Login	creates	UserSession
Attacker	exploits	SystemAccount System
SystActor	hasAction	UserAuth Login
BruteForce	hasConsequence	SystReact
SystActor	hasIdentification	Credentials

Login	hasPermission	TrueID Password
SystAccount	isIntegrated	System
Privilege	isPartOf	Role
Role	isPropertyOf	System
UserAuth	needs	Login
Attacker	teases	SystReact
Attacker	uses	Password Credentials Role Motivation
UserAuth	verifies	Credentials

Table 6 – The datatypes properties of the ontology #2

Domain	**Datatypes Properties**	**Range**
Attacker	guessPwd	String
Attacker	guessUserID	String
SystActor	hasIntention	Boolean
User	pwd	String
Attacker	stealPwd	String
Attacker	stealUserID	String
User	userID	String

With the brute force attack modelling and metrics evaluation approach it is possible to describe the development of the entities, objects properties, data properties and instances. The attacker can explore with detailed complexity, all intentions and goals. All components and techniques can be used in the plan to optimize the attack visualization. This expands the list of parameters to improve the model and allows the addition of important knowledge for the attack phases.

The ontology metrics presented in Table 7 were obtained from the Protégé ontology editor, version 5.5.0-beta-3. The difference between the metrics can be observed and compared in the ontology 1 column and the ontology 2 column.

In addition, the table shows the amount of classes, object properties, data properties and individuals of both classes. We can see that ontology #2 has more than 25 classes than ontology #1. The ontology design that can be observed in Figure 8 and in the Appendix 1 show clearly the dimensions of the attack scenarios. These additional categories include targets attacked, actors of the system, threats, brute force attack, user, attacker, motivation, mitigation, system account, privilege and roles that are considered.

In the evaluation between ontology #1 and ontology #2, the results are favorable to ontology #2, which has more information, it is more complex and it represents details between entities, classes and properties.

For the ontologies comparative analysis, it is possible to find redundancy where classes have the same formal definition. The representation of the classes and relationships is more specific and is not ambiguous when it allows a comprehensive view and a standardization of metrics.

Table 7 – Metrics with relevance

	Ontology metrics	**Ontology #1**	**Ontology #2**
Metrics	Axiom	113	164
	Logical axiom count	61	90
	Declaration axioms count	52	74
	Class count	17	42
	Object property count	8	15
	Data property count	5	7
	Individual count	22	10
	DL expressivity	AL(D)	AL(D)
Class axioms	SubClassOf	11	30
Object property axioms	ObjectPropertyDomain	12	15
Data property axioms	ObjectPropertyRange	8	21
	DataPropertyDomain	5	7
Individual Axioms	DataPropertyRange	5	7
	ClassAssertion	20	10

A set of experiments with the OWL ontologies allowed for topological analysis and evaluation of metrics of the brute force attack. The results show two attack scenarios: ontology #1 and ontology #2. After constructing the brute force attack graph, the Protégé provides the following information: metrics, knowledge, class axioms, object properties axioms, data property axioms and individual axioms.

For the first scenario, according to ontology #1 the knowledge about the attacker shows that the sequences are more restricted than in ontology #2. These results allow making decision about the most efficient countermeasures. These graphs demonstrate the main possibilities of the suggested evaluation scenario on metrics calculation. For the second scenario, there are several security mechanisms identified to mitigate specific threats. The alignment provides different calculation algorithms according to the available input data and allows to get adequate security assessment in any time of the system and the new data from the system influences on the probability and risk values of the attack.

The ability to infer class relationships enables a style of modeling in which subclass relationships are rarely asserted directly. Instead, relationships between classes are described in terms of unions, intersections, complements, and restrictions, and the inference engine determines the class structure. If more information is learned about a particular class or individual, then more class structure can be inferred [23].

Calculated metrics allow the determination of scenarios, including the existence of attacks, attacker skills and position and goals. The specifications of the structure of the security metrics include metrics, group of metrics and their interconnections and topological characteristics. The prediction of attacker steps is considered but the attacker skills and brute force attack criticality and potentiality. The risk impact is not considered in supporting the attack position.

One aspect that is observed is the number of relationships and there is a significant difference between the two ontologies. Figure 7 and appendix 1 is possible to see that ontology #2 is associated with more features. The mean number of properties per class means that probably the ontology is more useful.

In particular the metrics promotes the knowledge reasoning to infer in several cases related with threats and vulnerabilities.

4.2. Ontology quality improvement

The ontology quality metrics proposed in previous sections of this thesis, allows for an objective quantitative and objective analysis of ontologies quality. Following on the ontology metrics definition, we can look at ways of improving the quality of an ontology, taking into account the several quality criteria that ontologies reveal.

If we think about ontology quality attributes such as scope or knowledge domain coverage (number of concepts covered by an ontology), specialization (concepts hierarchy depth), operationality (concept instances present in an ontology), reasoning/logical inference ability (number of axioms and rules present in the ontology), etc. It is clear that we can measure objectively how interesting an ontology might be to represent knowledge of a certain domain, and how much potential it has to help on solving problems of that domain, using web semantic technologies and the corresponding computing tools.

In addition to utility based quality metrics, there are also ontology structure and operations metrics that might be taken into account, when we think of automatic processing of ontology by computation means.

The complex structure of the relationships between concepts represented in the ontology and its dynamic content during operation, requires the application of certain optimization procedures to improve ontology quality attributes such as response time to requests, and also adaption of its contents to the user/decision maker needs, by excluding those resources that are rarely or not used at all.

The identification and removal of ontology resources for the purpose of ontology normalization and the optimization of the ontology structure according to criteria such as access or reasoning performance, is considered of major importance to foster the use and utility of ontologies in current software systems and applications.

During the operation of information systems there is a constant filling of the ontology with new concepts that in turn, require periodic decision making regarding the selection of elements to be removed (graph reduction) to preserve the integrity of its semantic structure. In this regard, while making ontology changes there is a need for correction of the structure and content of the ontology.

Based on the characteristics of ISO/IEC 9126, the following optimality criteria of structure and content of the ontology must be considered:

- Physical memory occupied by the ontology;
- Speed of operation as response time of the information system to external requests;
- Completeness of the ontology, which can be determined using the average percentage of non-trivial (non-zero) responses to requests;
- Integrity of the ontology, that is the absence of mutually objecting claims and duplication;
- Balance of subject area, expressed as a uniform representation of its individual units in the ontology.

In order to optimize the ontologies that support information systems decision processes, it is necessary to choose more than one criterion or a combination of criteria to be optimized. The choice of the method for this combination of criteria has to be done based on experience and specific system requirements.

To a large extent, criteria are heuristic, they cannot be substantiated by something common to all systems mathematical positions. Additionally, the criteria to be optimized conflict with each other. For example, completeness, physical memory occupied and speed of access are typically conflicting objectives.

Most problems in nature, engineering, economy, industry, etc., have several, possibly conflicting, objectives to be satisfied, i.e. they are multiobjective optimization problems. Ontology quality optimization also fits in this type of optimization problems. We aim to optimize several ontology quality criterias simultaneously. Quality criteria such as knowledge domain coverage and ontology speed of access are somehow contradicting quality criteria. Multiobjective optimization algorithms are specially suitable for application in this kind of problems, to find the best trade-offs among several conflicting criteria.

Next we present a formal definition of a multiobjective optimization problem and describe the way ontology quality criteria optimization can be formulated as a multiobjective optimization problem.

A general optimization problem formulation is given by a quadruple (X, Z, fm, rel) where:

X is the decision space or input space. In decision space optimal solutions are known as efficient solutions/set. Elements of X are called decision vectors or simply solutions;

Z is the objective space. Each point in decision space maps to an m dimensional vector in the objective space containing the objective function values, which are compared to each other.

Optimal;

fm are the objective functions to optimize (minimize or maximize simultaneously). f: X→Z assign each decision vector an objective vector;

rel is a binary relation over Z, expressing preferences between solutions in the objective (quality) space.

A multiobjective optimization problem can be presented as a simultaneous optimization of m objective functions f=(f1,f2,...,fm), such that fk, k ∈ {1,...,m} are functions representing ontology quality criteria.

Figure 9 illustrates graphically the fundamental structure of a multiobjective optimization problem.

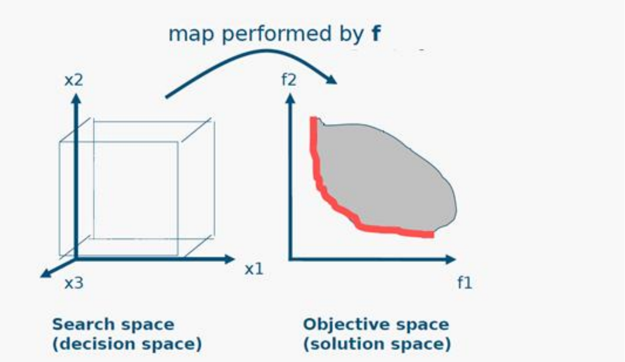


Figure 9 – Fundamental structure of a multiobjective optimization problem

The search space represents, in the case of ontology quality optimization, the decisions that can be made to change the ontology structure or content, e.g. removing concepts, instances, relations or axioms. These decisions have an effect in the objective (quality) space, measured (computed) by the defined objective functions (ontology quality metrics formulas) defined in previous sections of this thesis. The grey area in Figure 9 represents all possible solutions and the red line represents the best possible trade-offs for the formulated multiobjective optimization problem.

A variety of multiobjective optimization algorithms exist that can be applied to multiobjective optimization problems in general and to the ontology quality optimization problem in particular. One of the reference algorithms in multiobjective optimization that might be tried in the multiobjective optimization problem presented in this thesis is NSGA-II [64].

Along this thesis, the decision variables, objective functions and ontology quality criteria were presented in detail. In this section the problem of ontology quality improvement was presented as a multiobjective optimization problem. A reference multiobjective optimization algorithm was pointed out as of potential interest to be applied to the optimization problem identified in this thesis.

5. Conclusions and future work

5.1. Final summary

In this thesis the core importance of knowledge representation by the means of ontologies modelling and ontologies computational standards was studied and presented. Ontology based knowledge representation will facilitate the sharing, communication, reuse and harmonization of knowledge, allowing to build new levels of collaboration at the internet/web scale. One of the main challenges faced nowadays in the information society is the modeling of knowledge with so called semantic web standards and technologies to form integrated bodies of knowledge that can be understood and processed by humans or machines. This is progressively creating the grounds for a systematic and global knowledge sharing in all knowledge domains based in information coming from multiple sources in the cyberspace. The methodologies, how to build, maintain and take benefit of it in practice was covered in this thesis, by a set of recommendations and examples in the cybersecurity field.

The first goal of this thesis was to provide a synthesis and a comprehensive view of the role of ontologies in general and computer based ontologies in particular, introducing the perspective of ontologies quality metrics as the means to support sustainable knowledge sharing in the context of semantic web technologies. The potential existence of multiple ontologies on similar knowledge domains, raises the difficulty of deciding which ontologies best match our interests and quality criteria. An increase need of ontologies quality assessment was identified, since a urge amount of ontologies is being created recently in all knowledge domains, with different quality attributes, such as amount of asserted facts, domain knowledge scope covered, reasoning ability, etc. The need for computing systems to automatically analyze unstructured text and unstructured data from various data sources, at the web scale, turns ontologies and web semantic technologies into core building blocks of this transformation process.

The second goal of this study was to propose quality based metrics to examine differences between real world domain strategies that contain uncertain knowledge attributed to incomplete or partial information that is true only to a certain degree. However, because the two issues have some connection, it would be convenient to consider quality differences first. Clearly, there is much work to be done in current conceptualization and measurement of quality, in addition to what was presented in this thesis.

5.2. Contributions to the scientific and business community

The contribution of the study to be highlighted is related to the analysis and development of ontologies in cybersecurity. A complete cybersecurity ontology has not yet been accomplished by the scientific community and most of the work in cybersecurity ontologies has focused on specific domains. The goal of a complete ontology for the cybersecurity field cannot be an isolated task, since it is impossible to formalize all the existing concepts. It can only be achieved with the collaboration of all of the security community by joining and improving the developed ontologies for the specific domains. There are several ways the community can do this such as using a consistent model of best practice, greater international collaboration between organizations and better collaboration between the business and academic communities.

In order to illustrate the theory and practice of ontology design and engineering in the cybersecurity domain, a small ontology was designed and presented in this thesis, including the explanation of the design process difficulties, best practices and benefits.

5.3. Implications and limitations of research

There are some limitations of the present study that could be pointed out, such as the fact that the information collected to build the demonstration of cybersecurity ontology design, used only a few specific data sources. The interpretations and comprehension of the created ontology and the corresponding design process, can only be seen of interest and validity for research purposes and not for operations purposes.

5.4. Proposals for further research

OWL ontologies quality assessment, by the means of objective and quantitative metrics computation, was seen in this thesis as the basis for ontology quality attributes optimization in the perspective of a multiobjective optimization problem formulation.

A multiobjective optimization problem formulation was presented and described in detail, together with the identification of a suited algorithm to be applied to the formulated problem.

It is expected that the application and testing of this proposal on the udge amount of available cybersecurity OWL ontologies, will allow for a systematic quality assessment of the knowledge represented in this domain.

Bibliography

[1] ***Ontology.*** 2018. In https://en.wikipedia.org/wiki/Ontology (Accessed: 12/04/2018).

[2] Barchini, G., Álvarez, M., & Herrera, S. 2006. Sistemas de Información: Nuevos escenarios basados en ontologías. ***Journal of Information Systems and Technology Management***, 3(1): 3-18.

[3] Neches, R., Fikes, R. E., Finin, T., Gruber, T. R., Senator, T., & Swartout, W. R. 1991. Enabling technology for knowledge sharing. ***AI Magazine***, 12(3): 36-56.

[4] Gruber, T.R. 1993 A translation approach to portable ontology specification. ***Knowledge Acquisition***, 5(2): 199-200.

[5] Gruber, T.R. 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. ***International Journal Human-Computer Studies***, 43: 5-6.

[6] Smith, B. 2003. ***Ontology and Information Systems***, University at. Buffalo. In hppp://ontology.buffalo.edu/ontology%28PIC%29.pdf (Accessed 15/11/2017).

[7] Rees, R. 2003. ***Clarity in the usage of the terms ontology, taxonomy and classification***, In Proceeding of the CIB W78's 20th International Conference on Construction IT, Construction IT Bridging the Distance. 432-440. New Zealand: CIB Report 284.

[8] Malucelli, A. 2006. ***Ontology-based services for agents interoperability***, Faculdade de Engenharia. Department of Computing and Electrical Engineering. Universidade do Porto, Porto.

[9] W3C. 2012. ***OWL 2 Web Ontology Language.*** In https://www.w3.org/TR/owl2-overview/ (Accessed 25/11/2017).

[10] Borst, W.N. 1997. **Construction of Engineering Ontologies for Knowledge Sharing and Reuse**. Unpublished thesis, University of Tweenty, Enschede.

[11] Van Heist, G. 1997. Using explicit ontologies in KBS development. ***International Journal of Human-Computer Studies***, 47:183-292.

[12] Mizoguchi, R., Vanwelkenhuysen, J. Y., & Ikeda, M. 1995. Task ontology for reusable problem solving knowledge, Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing. ***IOS Press***, 46-59.

[13] Uschold, M., & Jasper, R. 1999. ***A Framework for Understanding and Classifying ontology Applications***, In V. R. Benjamins (Ed.), *IJCAI'99 Workshop on Ontology and Problem Solving Methods: Lessons Learned and Future Trends. 18: 11.1-11.12. Stockholm*: CEUR Workshop Proceedings. In http://CEUR-WS.org (Accessed 30/11/2017).

[14] Lévy, P. 1999. ***Cibercultura.*** São Paulo: Editora 34.

[15] Sheth, A.P. 1999. Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In M. Goodchild, Egenhofer, M.J., Fegeas, R., Kottman, C. (Eds.), ***Interoperating Geographic Information Systems***: 5-29. Boston: Springer.

[16] Baptista, A. 2002. ***Information Online Um Enquadramento para a Publicação em Linha de Revistas Científicas Eletrónicas***. Universidade do Minho, Guimarães.

[17] Souag, A., Salinesi, C., Mazo, R. & Comyn-Wattiau, I. 2015. A Security Ontology for Security Requirements Elicitation. In F.Piessens, J. Caballero, N. Bielova (Eds.), ***International Symposium on Engineering Secure Software and Systems***: 157-177. Switzerland: Springer International Publishing.

[18] Woods, W.A. 1975. What´s in a link: Foundations for semantic networks, In D.G. Bobrow & A.M. Collins (Eds.), ***Representation and Understanding: Studies in Cognitive Science****: 9-15*. New York: Academic Press.

[19] Sheth, A. 1995. ***Data semantics: What, where and how?*** .In 6th IFIP Working Conference on Data Semantics. DS-6. Georgia: Stone Mountain.

[20] Berners-Lee, T., Hendler, J., & Lassila, O. 2001. The Semantic Web. ***Scientific American***, 29-37.

[21] Berners-Lee, T. 2009. ***Linked Data***. In http://www.w3.org/DesignIssues/LinkedData.html (Accessed: 26/11/2017).

[22] Eckert, K. 2013. ***Provenance and Annotations for Linked Data.*** In International Conference on Dublin Core and Metadata Applications. 9-18, Dublin.

[23] Allemang, D., & Hendler, J. 2011. ***Semantic Web for the Working Ontologist Effective Modeling in RDFS and OWL*** (2nd ed.). New York: Morgan Kaufmann Publishers.

[24] Antoniou, G. & Harmelen, F.V. 2009. Web ontology language: OWL, In R.S. S. Saab (Ed.),.***Handbook on Ontologies****: 91-110.* Berlin: Springer-Verlab.

[25] Sattler, U., Calvanese, D. & Molitor, R. 2010. Relationships with other formalisms. In D.C. Franz Baader, Deborah L. McGuinness, Daniele Nardi, Peter F. Patel-Schneider (Eds.), ***The description logic handbook: Theory, implementation, and applications***: 142-183**.** New York: Cambridge University Press.

[26] Singhal, A., & Wijesekera, D. 2010. ***Ontologies for Modeling Enterprise Level Security Metrics***. In Proceedings of the 6th Cyber Security and Information Intelligence Research Workshop, CSIIRW 2010. Oak Ridge: ACM.

[27] Chowdhury, F., & Ferdous, Md S. 2017. Modelling Cyber Attacks. **International Journal of Network Security & Its Applications** (IJNSA), 9(4): 13-32.

[28] Syed, Z., Padia, A., Mathews, A., Finin, M. L., & Joshi, A. 2016. ***UCO: A Unified Cybersecurity Ontology.*** In Proceedings of the AAAI Workshop on Artificial Intelligence for Cyber Security: 195-202, University of Maryland, Baltimore County.

[29] Nunes, P.V. 2013. ***Cibersegurança e Estratégia Nacional de Informação: Estruturas de Coordenação Nacional no Ciberespaço***. In Conferência IV SIMSIC, Beja.

[30] Obrst, L., Chase, P. & Markeloff, R. 2012. Developing an Ontology of the Cyber Security Domain, In P.C.L. da Costa, Kathryn B., (Ed.), ***Semantic Technologies for Intelligence, defense, and Security (STIDS):*** 49-56, Fairfax: CEUR-WS.org.

[31] Silva, D.V. & D., Rodriguez. 2017. ***Ontologies for Network Security and Future Challenges.*** In Proceedings of the 12th International Conference on Cyber Warfare and Security - ICCWS 2017: 541-547, Dayton: Academic Publishing Limited.

[32] Arbanas, K., & Cubrilo, M. 2015. Ontology in Information Security. ***Journal of Information and Organizational Sciences***, 39(2): 107-136.

[33] Undercoffer, J., Josh, A. & Pinlston, J. 2003. Modeling computer attacks: An ontology for intrusion detection. In E.J. G. Vigna, and C. Kruegel (Eds.), ***International Workshop on Recent Advances in Intrusion Detection***: 113-135. Springer-Verlag.

[34] Undercoffer, J., Pinkston, J., Joshi, A. & Finin, T. 2004. ***A Target-Centric Ontology for Intrusion Detection***. In Proceeding of the IJCAI-03 Workshop on ontologies and distributed systems. Acapulco: Morgan Kaufmann Pub.

[35] Herzog, A., Shahmehri, N. & Duma, C. 2007. An Ontology of Information Security. ***International Journal of Information Security and Privacy (IJISP)***, 1(4): 1-23.

[36] Elahi, G., Yu, E. & Zannone, N.. 2009. ***A Modeling Ontology for Integrating Vulnerabilities into Security Requirements Conceptual Foundations***. In ER '09 Proceedings of the 28th International Conference on Conceptual Modeling. Gramado: Springer.

[37] Cuppens-Boulahia, N., Cuppens, F., Autrel, F. & Debar, H. 2009. ***An ontology-based approach to react to network attacks***, In Third International Conference on Risks and Security of Internet and Systems: 220-305, IEEE.

[38] Blackwell, C. 2010. ***A security ontology for incident analysis.*** In Sixth Annual Workshop on Cyber Security and Information Intelligence Research (CSIIRW '10), Oak Ridge.

[39] Pereira T., S.H. 2010. A Security Audit Framework to Manage Information System Security, In J.H. Tenreiro de Magalhães S., Hessami A.G. (Eds.), ***International Conference on Global Security, Safety, and Sustainability - ICGS3***, vol. 92: 9-18. Heidelberg: Springer.

[40] Souag A., S.C. & Comyn-Wattiau I. 2012. ***Ontologies for Security Requirements: A Literature Survey and Classification.*** In Advanced Information Systems Engineering Workshops. CAiSE. Lecture Notes in Business Information Processing. Gdansk: Springer.

[41] Gao, J.B., Zhang, B.W., Chen, X.H. & Luo, Z. 2013. Ontology-based model of network and computer attacks for security assessment. ***Journal of Shanghai Jiaotong University (Science)***, 18(5): 554-562.

[42] Kang, W. & LIang, Y. 2013. ***A Security Ontology with MDA for Software Development***. In International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery. Beijing: IEEE.

[43] Wali, A., Chun, S. A. & Geller, J. 2013. ***A Bootstrapping Approach for Developing a Cyber-security Ontology Using Textbook index Terms.*** In Proceedings International Conference on Availability, Reliability and Security (ARES 2013). Regensburg: IEEE.

[44] Ben-Asher, N., Oltramari, A., Erbacher, R.F. & Gonzalez, C. 2015 . ***Ontology-based Adaptative Systems of Cyber Defense.*** In Tenth Conference on Semantic Technology for Intelligence, Defense, and Security. Fairfax: CEUR-WS.org.

[45] Simmons, B.C., Shiva, S.G. & Simmons, L.L. 2014. ***A qualitative analysis of an ontology based issue resolution system for cyber attack management***. In The 4th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent: 323-329.

[46] Razzaq, A., Anwar, Z., Ahmad, H.F., Latif, K. & Munir, F. 2014. Ontology for attack detection: An intelligent approach to web application security. ***Computers & Security***, vol. 45: 124-146.

[47] Sartakov, V.A. 2015. Ontological Representation of Networks for IDS in Cyber-Physical Systems. In K.N. Khachay M., Panchenko A., Ignatov D., Labunets V. (Eds.). ***Analysis of Images, Social Networks and Texts.*** AIST 2015.Communications in Computer and Information Science, *vol. 542: 421-430*, Cham: Springer.

[48] Kyriakopoulos, K.G., Parish, D.J. & Whitley, J.N. 2015. ***FlowStats: An ontology based network management tool***, In Second International Conference on Computing Technology and Information Management (ICCTIM), Johor: IEEE.

[49] Lannacone, M., Bohn, S., Nakamura, G., Gerth, J., Huffer, K., Bridges, R. & Goodall, J. 2015. ***Developing an ontology for cyber security knowledge graphs.*** In Proceedings of the 10th Annual Cyber and Information Security Research Conference. Oak Ridge: ACM.

[50] Salem, M.B. & Wacek, C. 2015. ***Enabling New Technologies for CyberSecurity Defense with the ICAS Cyber Security Ontology***. In Semantic Technology for Intelligence, Defense, and Security (STIDS 2015) George Mason University, Fairfax.

[51] Costa, D., Albrethsen, M., Collins, M., Perl, S., Soliwash, G. & Spooner, D. 2016. ***An Insider Threat Indicator Ontology.*** Pittsburgh: Carnegie Mellon University.

[52] Roldán-Molina, G., Almache-Cueva, M., Yevseyeva, I., Silva-Rabadão, C., & Basto-Fernandes, V. 2017. ***A decision support system for corporations cybersecurity management.*** In 12th Iberian Conference on Information Systems and Technologies (CISTI). IEEE, Lisbon.

[53] Martín, L.T. 2018. ***Ontologías en la gestión de redes basada en políticas.*** In 3rd International Workshop on Semantic Web (IWSW 2018). CEUR-WS, Havana.

[54] Bojanova, I.B., P.E., Yesha, Y. & Wu, Y. 2016. ***The Bugs Framework (BF): A Structured Approach to Express Bugs.*** In IEEE International Conference on Software Quality, Reliability and Security (QRS). IEEE, Vienna.

[55] Mitchell, W. 2018. ***Password cracking.*** In http://web.cs.du.edu/~mitchell/forensics/information/pass_crack.html (Accessed 03/09/2018).

[56] Oltramari, A., Cranor, L. F., Walls, R. J., & MacDaniel, P. D. 2014. ***Building an ontology of cyber security.*** In Proceedings of the Ninth Conference on Semantic Technology for Intelligence, Defense, and Security: 54-61, Fairfax.

[57] Hevner, A., March, S., Park, J., & Ram, S., 2004. Design science in information systems research. ***MIS Quarterly***, 28(1): 75-105.

[58] Mohsen, W., Aref, M. & Elbahnasy, K. 2017. ***Software metrics for cooperative scrum based ontology analysis***. In 2nd International Conference on Knowledge Engineering and Applications (ICKEA): 60-70, IEEE Xplore.

[59] Cimiano, P.V., J. 2005. ***Text2Onto - A Framework for Ontology Learning and Data-driven Change Discovery.*** In NLDB'05 Proceedings of the 10th international conference on Natural Language Processing and Information Systems, Heidelberg: Springer-Verlag.

[60] Ramos, A.D. 2016. ***Framework basado en el estándar de calidad del software ISO/IEC 25000:2005 (SQuaRE) para la evaluacíon de la calidad de las ontológias.*** Unpublished dissertation. Universidad de Murcia, Murcia.

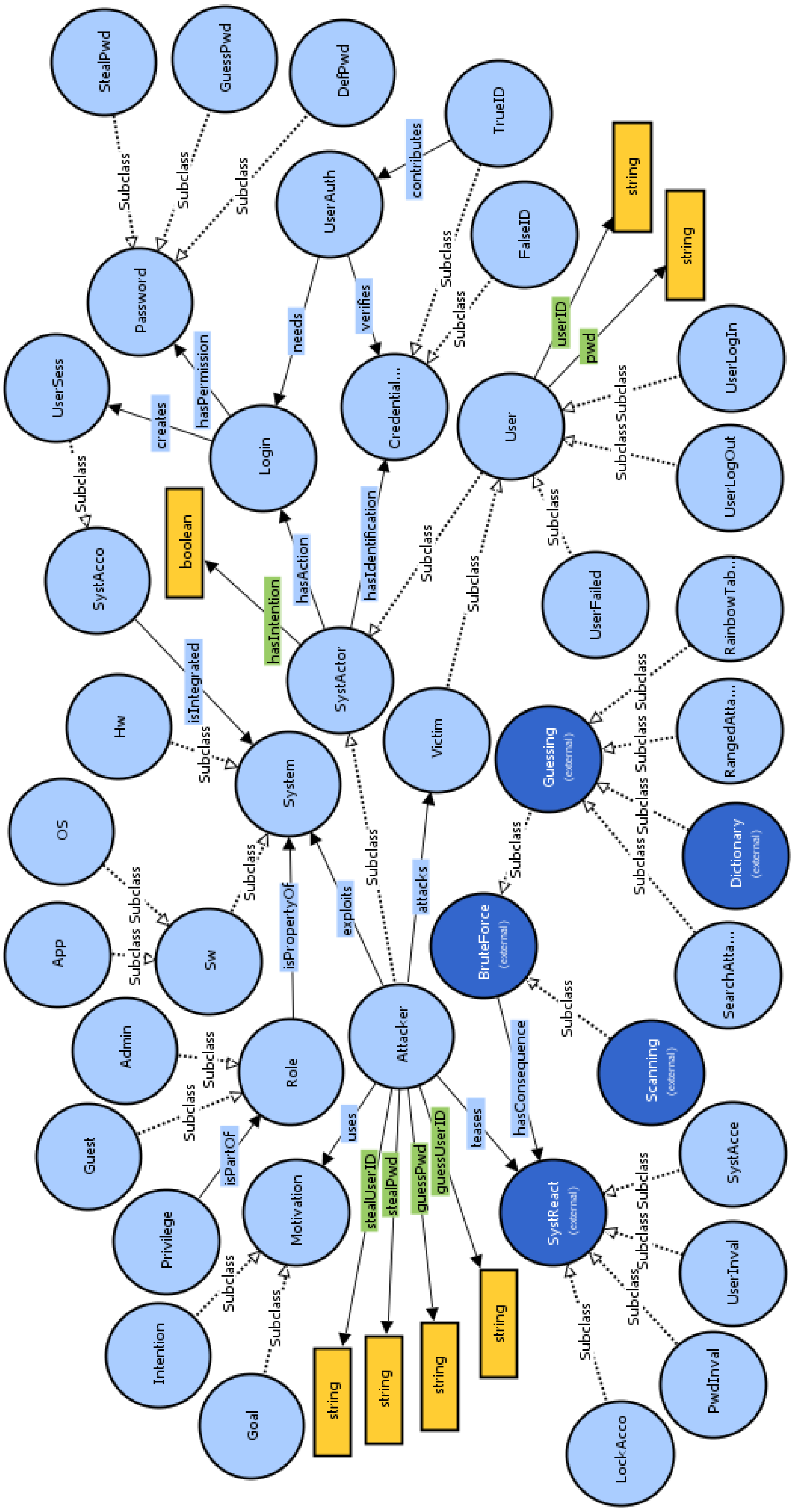
[61] Khan, S.A.Q., M.A., Abbas, M.A. & Afzal, M. T. 2017. OWL2 benchmarking for the evaluation of knowledge based systems. ***Plos one***, 12(6): e0179578.

[62] Noy, N., McGuinness, D. 2001. ***Ontology Development 101: A Guide to Creating Your First Ontology***. Stanford: Knowledge Systems Laboratory.

[63] Hnatkowska, B.W.P. 2018. Transformation of OWL2 property axioms to Groovy. In Tjoa A., Bellatreche L., Biffl S., van Leeuwen J., Wiedermann J. (Eds.), ***SOFSEM 2018: Theory and Practice of Computer Science***, vol 10706: 269-282. Cham: Springer.

[64] Deb, K., Pratap, A., Agrawal, S., & Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: in NSGA-II. ***IEEE Transactions on Evolutionary Computation***, 6(2): 182-197.

Appendix 1: The ontology #2 graph representation


StealPwd
GuessPwd
DefPwd
TrueID
UserAuth
contributes
FalseID
string
string
Password
Subclass
UserSess
creates
hasPermission
needs
verifies
Login
Credential...
User
userID
pwd
UserLogIn
UserLogOut
SystAcco
boolean
hasAction
hasIdentification
hasIntention
SystActor
UserFailed
RainbowTab...
Hw
isIntegrated
System
Victim
Guessing
(external)
RangedAtta...
OS
Dictionary
(external)
App
Sw
isPropertyOf
exploits
attacks
BruteForce
(external)
SearchAtta...
Admin
Role
Attacker
hasConsequence
Scanning
(external)
Guest
isPartOf
uses
teases
SystAcce
Privilege
Motivation
stealUserID
stealPwd
guessPwd
guessUserID
SystReact
(external)
UserInval
Intention
Goal
string
string
string
string
PwdInval
LockAcco